

STATISTICAL DEFORMABLE MODEL APPLIED TO ANATOMICAL LANDMARK DETECTION

Camille Izard, Bruno Jedynek

Laboratoire Paul Painlevé, Université des Sciences et Technologies de Lille, Villeneuve d'Ascq, France
Center for Imaging Science, Johns Hopkins University, Baltimore MD

ABSTRACT

We present a generic statistical deformable model for gray level medical images and propose to use it for template matching. Template matching methods usually rely on the arbitrary choice of a cost function and a template. Statistical models, on the other hand, allow us to derive optimal learning and matching algorithms from the modeling assumptions using likelihood maximization principles. We test the statistical deformable model on the automatic anatomical landmark detection in brain MRI, and compare its performance with the sum of squared differences (SSD), a reference cost-function for intensity-based template matching.

Index Terms— statistical models, deformable template, anatomical landmarks

1. INTRODUCTION

Deformable template models (see e.g. [1]) have been broadly used in numerous applications in medical imaging, such as image segmentation, image registration and shape analysis. The matching of the template onto an image is generally written as the optimization of an energy function composed of the sum of two weighted terms: the data attachment and the regularization terms. Many energy functions resulting from different choices have been proposed [2]. While some of them perform well, the need of tailoring the cost function to the problem prevents from developing generic deformable models. However, when the images are modeled by a statistical deformable model [3, 4, 5, 6], it is possible to derive from the modeling assumptions both learning and matching algorithms using generic principles such as likelihood maximization. In this paper we propose a statistical deformable model of the image intensity and show how it can be used to derive intuitive yet mathematically sound algorithms to solve the problem of automatic landmark detection. We compare the resulting algorithm with the well-known sum of squared differences (SSD), a reference cost function for intensity-based matching [7]. Performance is assessed on the detection of four landmarks on two sets of 2D brain MRI.

Sponsored by the Doctoral Fellowship of the Université des Sciences et Technologies de Lille.

2. INTENSITY DEFORMABLE MODEL

We denote by Λ a finite lattice of \mathbb{R}^d ($d = 2$ or 3). At each pixel $s \in \Lambda$ a real value $x(s)$ is observed. $x \in \mathbb{R}^S$ is the vector of image intensity with S the number of pixels. We denote by $y \in \mathbb{R}^{dK}$, a set of K landmark locations in an image of dimension d . Both the image x and the landmark locations y are modeled as continuous random vectors. Let us further denote by x_0 a template image defined on a finite grid of pixels Λ_T , in which the landmarks lie in a fixed reference configuration \bar{y} . In the deformable model setting an image is assumed to result from the action of a random deformation on a template image x_0 . The deformation maps the reference landmark location \bar{y} to its position y in the image. We denote by $f_{y^{(i)}}$ a smooth deformation that maps the template grid Λ_T onto the i th image grid Λ_i , such that: $f_{y^{(i)}}(\bar{y}) = y^{(i)}$.

2.1. Deformation Model

We use a spline-based deformation model in which the deformation is parameterized by the displacement of the landmarks. It is extended to the rest of the image support by spline interpolation.

$$\forall t \in \Lambda_T, \quad f_y(t) = \sum_{k=1}^K \kappa(t, \bar{y}_k) \beta_k, \quad (1)$$

with $\beta_k \in \mathbb{R}^d$. For a given kernel κ , under mild conditions, there exists a unique deformation $f_y(t)$ that matches the reference locations \bar{y} to a specific landmark configuration y in the image. We choose κ to be a Gaussian kernel because of its simple analytical expression and because of its local support. When both the kernel and the reference landmarks are fixed, the landmark localization problem can be seen as a template matching problem, i.e. it is equivalent to find the location of the landmarks or to find the transformation that maps the template to the image.

2.2. Deformable Intensity Model

The Deformable Intensity Model (DIM) encodes the joint probability of the image intensities and of the landmark location $p(x, y)$. We choose an flat prior $p(y)$, which assigns

the same probability to all locations in the image, hence it is enough to work on the conditional distribution $p(x|y)$. As it is often the case in generative models of images, we will assume statistical independence of the image intensities given the location of the landmarks. Thus, the conditional probability can be written as a product over the image pixels. We assume that the intensity $x(s)$ follows a Gaussian distribution whose parameters $(x_0(f_y^{-1}(s)), \tau^2(f_y^{-1}(s)))$ depend on s and on the location of the landmarks, $\forall s \in \Lambda$,

$$x(s) = x_0(f_y^{-1}(s)) + \epsilon(s), \quad \epsilon(s) \sim \mathcal{N}(0, \tau^2(f_y^{-1}(s))). \quad (2)$$

Instead of a template image x_0 , we introduce a probabilistic template, which assigns to each location $t \in \Lambda_T$ a Gaussian distribution of parameters $(x_0(t), \tau^2(t))$. We denote $\pi(u, t)$ the probability of observing the intensity u at the pixel $t \in \Lambda_T$. The log-likelihood of an image x , given that the landmarks lie in y , has the following form:

$$\begin{aligned} \ell(x|y; x_0, \tau^2) &= \sum_{s \in \Lambda} \ln \pi(x(s), f_y^{-1}(s)) \\ &= - \sum_{s \in \Lambda} \left[\frac{(x(s) - x_0(f_y^{-1}(s)))^2}{2\tau^2(f_y^{-1}(s))} + \ln \tau(f_y^{-1}(s)) \right]. \end{aligned} \quad (3)$$

It is the same cost function as SSD, except that the variance τ depends on the location in the template, and therefore on the deformation f_y parameterized by the location of the landmarks.

3. MODEL SELECTION USING A TRAINING SET

3.1. Maximum Likelihood Estimation

We denote by $x_1^N \in (\mathbb{R}^S)^N$ the training set of N images, and by $y_1^N \in (\mathbb{R}^{dK})^N$ the location of the landmarks in the training images. We denote by x an image of the testing set. First, the model parameters, i.e. $(x_0(t), \tau^2(t)), \forall t \in \Lambda_T$, need to be estimated. We maximize the likelihood of the training set with respect to the model parameters:

$$(\hat{x}_0, \hat{\tau}^2) = \arg \max_{x_0, \tau^2} \ell(x_1^N | y_1^N; x_0, \tau^2). \quad (4)$$

Then, for a new image x , the location of the landmarks is predicted by the maximum likelihood estimate of y given the model parameters learnt during training:

$$\hat{y} = \arg \max_y \ell(x|y; \hat{x}_0, \hat{\tau}^2). \quad (5)$$

3.2. Direct Estimation of the Deformable Model

Two issues arise when trying to maximize (4). Depending on the training image i , different locations in the template $f_{y(i)}^{-1}(s) \in \Lambda_T$ correspond to a fixed pixel $s \in \Lambda$. Therefore the training images need to be registered before learning.

However, even though we assume that the inverse of $f_y^{(i)}$ exists for all image i , its inverse does not need to admit a simple analytical form. To avoid computing the inverse deformation and to overcome the estimation issue, we propose to approximate the likelihood function by a change of variable. First, the sum over all the pixels of the image support is approximated by the integral over the same image support. Then, for each image i , we perform the change of variable $s = f_{y(i)}(t)$, and denote by $|J_{f_{y(i)}}(t)|$ the absolute value of the deformation Jacobian at t . The likelihood $\ell(x_1^N | y_1^N; x_0, \tau^2)$ becomes:

$$\sum_{i=1}^N \sum_{s \in \Lambda_i} \ln \pi(x^{(i)}(s), f_{y(i)}^{-1}(s)) \quad (6)$$

$$\approx \sum_{i=1}^N \int_{f_{y(i)}^{-1}(\Lambda_i)} \ln \pi(x^{(i)}(f_{y(i)}(t)), t) |J_{f_{y(i)}}(t)| dt, \quad (7)$$

$$\approx \sum_{i=1}^N \sum_{t \in \Lambda_T} \ln \pi(x^{(i)}(f_{y(i)}(t)), t) |J_{f_{y(i)}}(t)|. \quad (8)$$

Finally, each integral is discretized and approximated by a sum over the template grid locations $t \in \Lambda_T$. In consequence it is possible to change the order of the sums and simplify the joint optimization. The maximization is performed independently at each pixel of Λ_T . The parameter estimates can be written in closed form: $\forall t \in \Lambda_T$,

$$\hat{x}_0(t) = \frac{\sum_{i=1}^N x(f_{y(i)}(t)) |J_{f_{y(i)}}(t)|}{\sum_{i=1}^N |J_{f_{y(i)}}(t)|}, \quad (9)$$

$$\hat{\tau}^2(t) = \frac{\sum_{i=1}^N [x(f_{y(i)}(t)) - \hat{x}_0(t)]^2 |J_{f_{y(i)}}(t)|}{\sum_{i=1}^N |J_{f_{y(i)}}(t)|}. \quad (10)$$

Note that the Jacobian weights the intensity samples coming from the training images. The template estimation consists in registering the training images based on their landmark correspondences, and averaging the intensity values, weighted by the Jacobian value. Because the deformations have a local support and because the landmarks are sparsely distributed in the image, the registration is expected to be better around the landmarks than at further distance. It means that the intensity variance τ^2 will be lower around the landmarks than at further distance, as depicted in Figure 1.

4. LANDMARK DETECTION

We use likelihood maximization to estimate the location of the landmarks y in a new image x , using the parameters learnt during training $(\hat{x}_0, \hat{\tau})$.

4.1. Optimization Method

The optimization is performed by a steepest gradient ascent combined with a line search method to optimize the step size at each iteration.

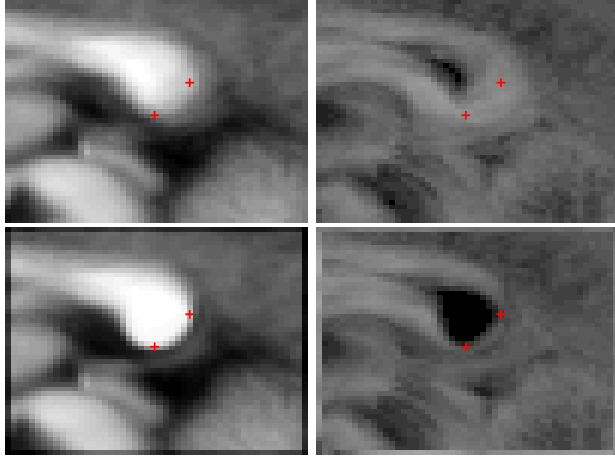


Fig. 1. Probabilistic Template. Left: average intensity, Right: standard deviation of the intensity. The top line represents the average intensity image, i.e. without registering the images first. The second line corresponds to the estimated template, i.e. after landmark-based registration of the training images. The red crosses represent \bar{y} .

1. Initialize the gradient ascent with $y \leftarrow \bar{y}$,
2. Iterate until convergence:
 - (a) Compute $\nabla_y \ell(x|y; \hat{x}_0, \hat{\tau}^2)$,
 - (b) $a \leftarrow \arg \max_{a \geq 0} \ell(x|y + a \nabla_y \ell(x|y; \hat{x}_0, \hat{\tau}^2); \hat{x}_0, \hat{\tau}^2)$,
 - (c) $y \leftarrow y + a \nabla_y \ell(x|y; \hat{x}_0, \hat{\tau}^2)$.

To avoid computing the inverse f_y^{-1} in (3), we approximate the likelihood as we did in section 3.2. It only affects the numerical computation of the likelihood and its gradient. We validated numerically this approximation on synthetic images, cf. [8].

4.2. Local Intensity Matching for Landmark Detection

Notice that the likelihood increases if the intensities of the deformed template matches the intensities of the image. When using SSD, the noise parameter τ which corresponds to the variance of the Gaussian distribution model, is constant throughout the template, such that all pixels have the same weight. Because the variance in DIM varies depending on the location in the template, for a same intensity difference, the pixels with lower variance contribute more to the cost function. Since the pixels surrounding the landmarks have lower variance, they have a greater weight in the likelihood variations.

Similarly to matching methods based on SSD, if the intensity distribution in the image differs significantly from the template intensity distribution, the best match in terms of likelihood may not correspond to the geometric match.

5. DETECTION RESULTS

5.1. Description of the Images

We use 47 T1-weighted Magnetic Resonance (MR) brain images acquired on a Philips-Intera 3-Tesla scanner, with resolution 1mm^3 , encoded in gray-level intensity between 0 and 255 after normalization. Brains were first manually transformed into standardized Talairach space to provide a canonical orientation and approximate alignment. In that orientation a specialist located manually some anatomical landmarks. The first data set is made of the mid-sagittal plane and contains the splenium of the Corpus Callosum. The splenium tip (SCC1) is defined as the most posterior extent of the corpus callosum and SCC2 is defined as the lower extent of the Splenium of the Corpus Callosum. The second data set is made of the sagittal slice that contains the Head of the Hippocampus (HoH), defined as the furthest extent of the hippocampus (in anterior and inferior direction). The main challenge to locate HoH is the absence of a visible contour around the hippocampus head. The Hippocampus Tail (HT) is defined as the furthest extent of the hippocampus (in posterior and superior directions) on the same sagittal slice. 17 images are sampled randomly from the data set and kept aside for testing. The remaining 30 images are used for training.

5.2. Detection Results in brain MRI

We train the model with $\sigma = 7$ pixels. The learnt template is represented in Figure 1. After training, landmarks are detected on the 17 test images. The Euclidean distance between manual landmarks and estimated landmarks measures the performance of the algorithm. We compare the performance of DIM and SSD, using the same deformation model and a gradient method. In addition we test the effect of neglecting the Jacobian for the estimation of the template. We denote these experiments DIM-A and SSD-A. "Initial" represents the distribution of the landmarks before detection.

Figure 2 presents the performance of the 5 predictors on the simultaneous detection of SCC1 and SCC2. All estimators produce better localization results than "Initial". We use a Wilcoxon test to assess the difference between the estimator performance. For SCC1 comparing DIM and SSD, we obtain a p-value of 0.0850. The p-value comparing DIM and DIM-A is 0.8904. As for SCC2, SSD and DIM perform equally.

Figure 3 presents the spatial repartition of the detection error of DIM around the real landmark locations. Notice that the error is locally aligned with the contour of the structure.

We study the effect of the kernel parameter σ on the detection performance. We vary the spline kernel standard deviation between 3 and 15 pixels. For larger standard deviation, the displacement of the landmarks influences a larger amount of pixels around the landmarks and therefore the size of the discriminative pattern increases.

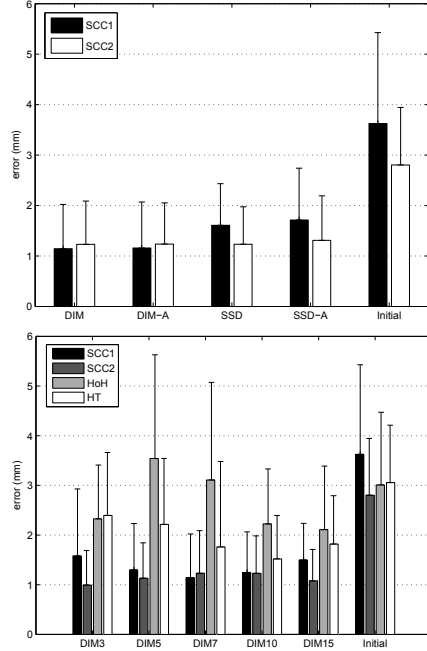


Fig. 2. Top: Performance of the 5 estimators for the prediction of SCC1 and SCC2. Bottom: Localization error and kernel parameter choice for 4 anatomical landmarks in the case of DIM with variable kernel standard deviation.

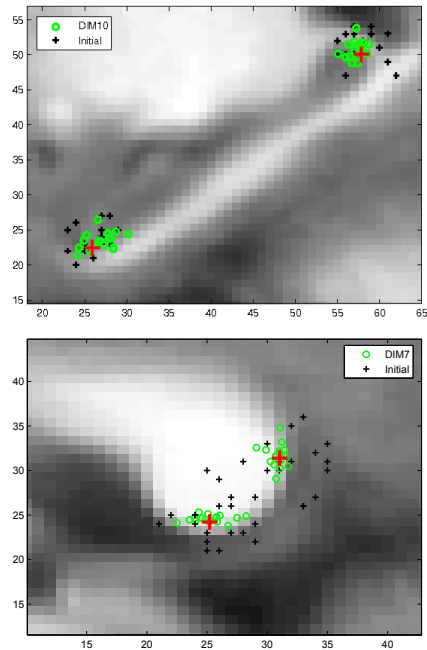


Fig. 3. Spatial representation of the detection error. The background image represents the template intensity average x_0 . The red crosses correspond to \bar{y} , the black crosses to the initial distribution of the localization error and the green circles to the residual error after detection.

Figure 2 also presents the performance of DIM with different kernel parameter values. The best performance is achieved for a standard deviation of 10 pixels. Depending on the landmark, the detection performance is more or less sensitive to the choice of the kernel parameter. The detection of HoH, for example, is quite good for DIM10. It deteriorates though in the case of DIM5 because the discriminative pattern is not large enough and slides along the white-gray boundary.

6. CONCLUSION

We have shown how the proposed statistical model can be used to derive template estimation and image matching algorithms. The method adapts to a variable number of landmarks, which means that it could be extended to other applications such as registration or segmentation.

The model we presented could be made more accurate by adding hidden variables and/or parameters. One can expect the estimation algorithm to become more complicated. Similarly the model could deal with other features of the image such as edges or even tissue type. However the main motivation of this paper is to emphasize that statistical models make it possible to derive unified algorithms and to relate modeling assumptions to changes in the algorithm, avoiding the arbitrary choice of a cost function.

7. REFERENCES

- [1] A. K. Jain, Y. Zhong, and M. P. Dubuisson-Jolly, "Deformable template models: A review," *Sign. Proc.*, vol. 71, pp. 109–129, 1998.
- [2] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image Vision Comput.*, vol. 21, pp. 977–1000, 2003.
- [3] A. Roche, G. Malandain, and N. Ayache, "Unifying maximum likelihood approaches in medical image registration," *Int. J. Imag. Syst. Tech.*, vol. 11, pp. 71–80, 2000.
- [4] C.A. Glasbey and K.V. Mardia, "A penalized likelihood approach to image warping (with discussion)," *J. Roy. Stat. Soc.*, vol. B, no. 63, pp. 465–514, 2001.
- [5] T.F. Cootes and C.J. Taylor, "Statistical models of appearance for medical image analysis and computer vision," in *SPIE*, 2001, vol. 4322, pp. 236–248.
- [6] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, pp. 839–851, 2005.
- [7] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital image registration," *IEEE Trans. Computers*, vol. 21, no. 2, pp. 179–186, 1972.
- [8] C. Izard, *Statistical Modeling and Estimation in Medical Imaging: Automatic Detection of Anatomical Landmarks*, Ph.D. thesis, Université des Sciences et Technologies de Lille, May 2008.